an abundance of at least 5% relative to all Core 1, Core 2, Core 3, Core 4, Core 5, Core 6, Core 7, and Core 8 O-glycans.

[0013] In an embodiment, the polypeptides expressed by the cells as described above include a transmembrane anchor that comprises a cytoplasmic recycling motif.

[0014] Isolated polynucleotides that encode the polypeptides of this disclosure are included, as are expression vectors comprising such polynucleotides. In embodiments, the polynucleotides are incorporated into the modified cells such that they are integrated into a chromosome of the cells, which may be achieved by random integration.

[0015] The disclosure includes a method of making cells that express the described polypeptides. This approach comprises introducing an isolated/recombinant polynucleotide that encodes a described polypeptide into the cells such that the polypeptide is expressed.

[0016] Also included is a method for producing a desired polypeptide or another agent, which may be distinct from the polypeptide comprising the repeated sequences. The method comprises expressing the desired polypeptide or producing another agent in modified mammalian cells that express a modified mucin polypeptide described herein. The method may further comprise separating the desired polypeptide or the other agent from the cells.

BRIEF DESCRIPTION OF THE FIGURES

[0017] The figures and tables of this disclosure are divided into four Parts (Part I, Part II, Part III, and Part IV), as described below.

Part I Figures

[0018] FIG. **1**: Combinatorial Genetic Encoded Library for Sequence-Specific Mucins. (a) Schematic diagram of the combinatorial sequence-specific mucins. (b) Schematic shows the swappable bio-bricks and flanking restriction sites for complete mucin construction. (c) Work flow for the design and fabrication of cDNAs for the mucin tandem-repeat backbones. (d) Summary of codon-scrambled mucin backbones in the library. The Wild-type Muc1 sequence is SEQ ID NO:8. The Muc1 single mutant (Muc1_S) is SEQ ID NO:5. The Muc1 double mutant (Muc1_D) is SEQ ID NO:6. The Muc1 triple mutant (Muc1_T) is SEQ ID NO:7. The Synthetic 1 (Syn1) is DAATPAP is SEQ ID NO:2. The Synthetic 2 (Syn2) is SEQ ID NO:3. The Synthetic 3 (Syn3) is SEQ ID NO:4. The Lubricin consensus sequence (Syn4) is SEQ ID NO:1.

[0019] FIG. **2**: Construction and Validation of Sequence-Specific Mucin Expression. (a) Components and features of codon-optimized Muc1 variants with GFP reporters. The amino acid sequence in (a) is SEQ ID NO:8. (b) Predicted Molecular Weight of the polypeptide backbone. (c) Biosynthesis of Tn antigen, Core 1, and Core 2 glycans, and specificity of relevant lectins for their detection. (d)Western Blot analysis of Native Muc1 expression and glycosylation in wild-type and Core-1 β3-T specific molecular chaperone (COSMC) knockout MCF10A cells. The MCF10A cells were stably transfected with native Muc1. The surface sialic acids were labeled with AFDye 568 through periodate labeling prior to lysate collection. The blot was stained in multiple colors with MUC1 TR (CD227 HPMV) Ab-FITC, and PNA-CF640 or biotinylated VVA (Secondary: NeutrAvidin-Dylight 650). (e) Western blot analysis of native and codon-scrambled Muc1 in extracts of transiently trnsfected HEK293T cells. (f) Immunofluorescence images of transiently transfected HEK293T cells expressing indicated constructs and probed with PNA lectin (left), anti-Muc antibody (center left), GFP (center right) and Hoescht nuclear stain (right) (scale bar 10 m). (g) PNA lectin blot analysis (left) and intensity profiles (right) of mucins of varying sizes in extracts of transiently transfected HEK293T cells.

[0020] FIG. **3**: Engineering the Frequency of Glycosylation Sites in the Muc1 Polymer Backbone Tunes O-glycan Maturation. (a) Components and features of secreted Muc1 and engineered variants each with 21 tandem repeats. (b) Tandem repeat sequences of secreted mucin mutants and the molecular weight of the polypeptide backbones. Single, double, and triple glycosylation mutants (sMuc1S, sMuc1D, and sMuc1T) have one, two or three, serine/threonine (S/T) to alanine substitutions per repeat, respectively. The sequences under sMuc1 mutants (21 repeats) are from top down: SEQ ID NO:8, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7. (c) Representative Western blot analysis of affinity-purified recombinant secreted mucins from FreeStyle™ 293-F cell culture media probed with anti-SUMOstar antibody and PNA, s-WGA and VVA lectins (of three independent experiments). The lectin blot was co-stained in multiple colors with PNA-Alexa Fluor 568, s-WGA-FITC, and biotinylated VVA (Secondary: NeutrAvidin-Dylight 650). (d) Representative fluorescence intensity electrophoretograms of the blots in (c). (e) Ratiometric intensity analysis of PNA to VVA signal (upper) and s-WGA to VVA signal (lower) for the indicated mucins and their corresponding frequency of S/T glycosylation sites in the polymer backbone. Ratiometric fluorescence intensity was quantified along each lane and normalized to signal from the secreted mucin with wild-type Muc1 tandem repeats (sMuc1); data presented as the mean and SEM from at least three independent experiments. * $P<0.05$ ** $P<0.01$ *** $P<0.001$ (f) Left: MALDI-TOF mass spectra registered for samples of permethylated glycan alditols from secreted mucins with wild-type Muc1 tandem repeats (sMuc1) and triple mutant (sMuc1T) from HEK293T cell culture media. The ion signals were annotated with respect to the relative masses of molecular ions (m/z) detected as sodium adducts and by assignment of the respective core structure (red for Core 1 and black for Core 2). Right: Schematic presentation of O-linked glycans detected on the secreted mucins.

[0021] FIG. **4**: Designer Mucin Domains Reveal Sequence-Specific Effects on Glycosylation. The sequences shown in FIG. **4** are KEPAPTTP (SEQ ID NO:1) DAATPAP (SEQ ID NO:2) DAATPAPP (SEQ ID NO:3) and PASTSAPG (SEQ ID NO:4). (a) Components and features of designer mucins. (b) Predicted Molecular Weight of the mucin polypeptide backbones. (c) Representative Western blot analysis (from three independent experiments) of indicated constructs in extracts of transiently transfected HEK293T cells probed with anti-GFP antibody or co-stained with PNA and VVA lectins. (d) Representative Fluorescence intensity electrophoretograms of the western blots in (c) for indicated constructs from three independent experiments. Dashed lines indicate the peak of the glycoform visible in the PNA blot. Shaded boxes indicate the regions between the bands on the anti-GFP blot with the highest and second highest apparent molecular weights. (e) Ratiometric intensity analysis of PNA to VVA staining for the indicated mucins and their corresponding frequency of serine and